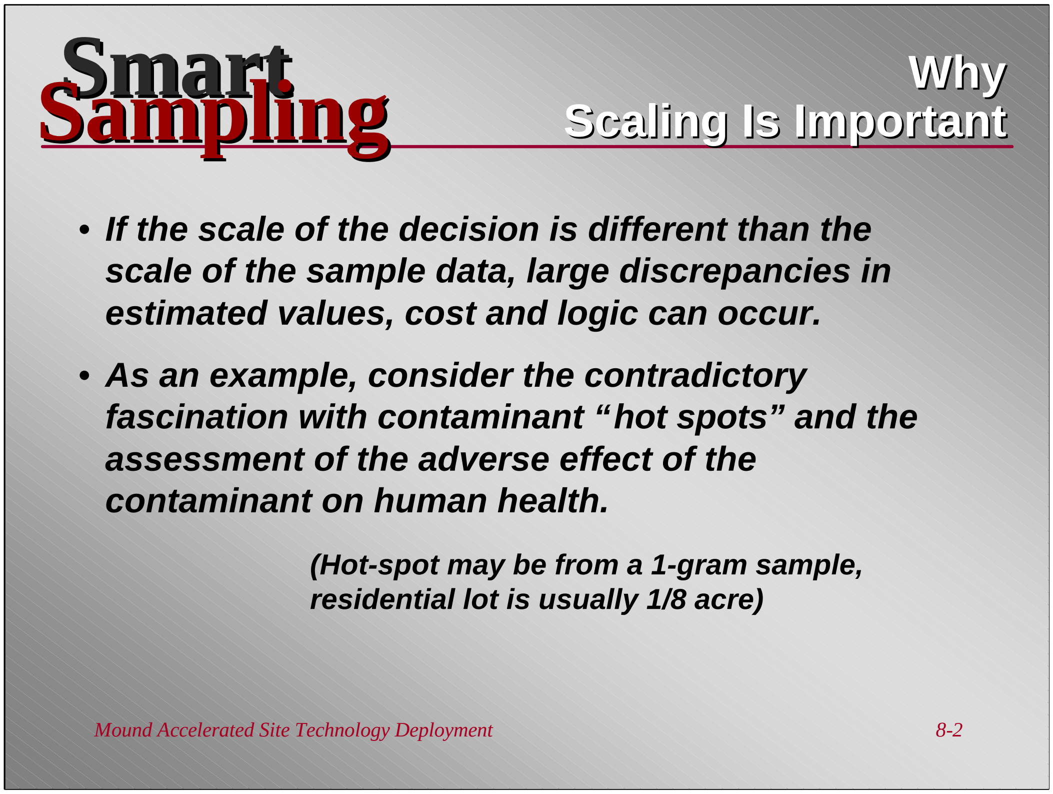
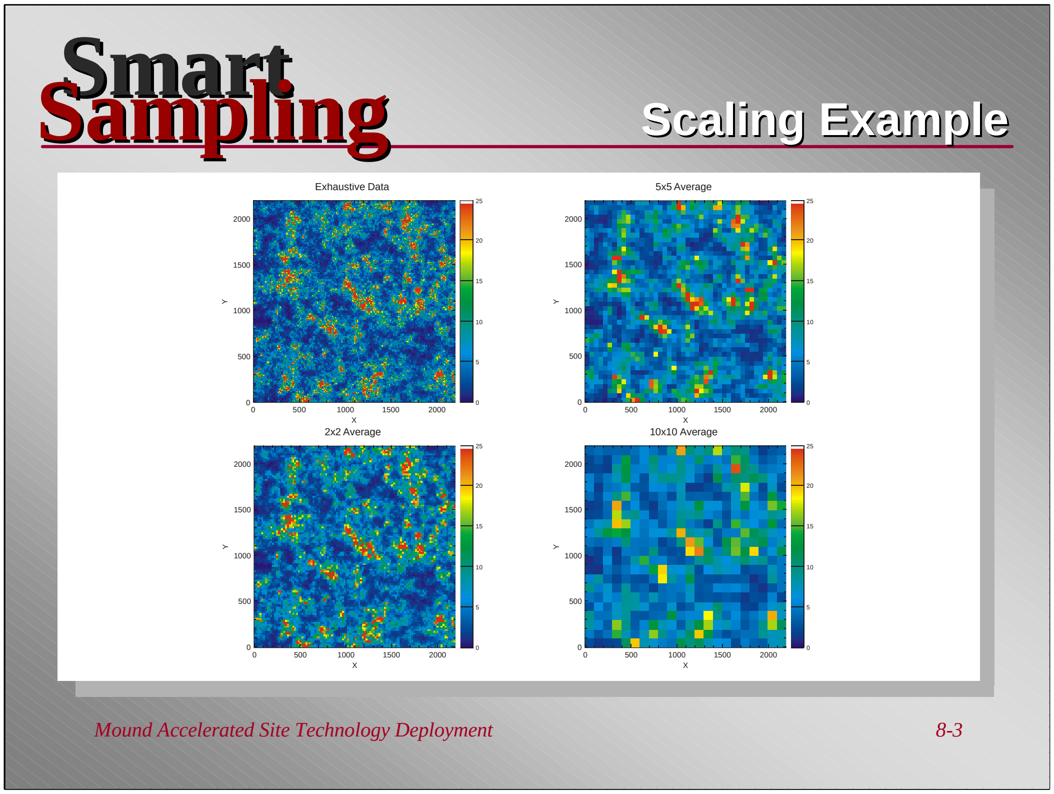
Section 8

Scaling



We have been looking at data sets that are well behaved in terms of samples being taken over the same scale that the decision is being made. That is not always the case. Often, or even usually, measurements from a sample size of a few centimeters of soil are used to make a decision over a 50 or 100 sq. ft. area. We don’t have samples and decisions at the same scale.

This section addresses how we mediate between those scales.



The site in this example is 2200 m x 2200 m. Composite samples were taken on 10 x 10m blocks. The concentrations were averaged to make a decision on 20 x 20m or larger blocks.

As the decision blocks get larger to 50 x 50m and 100 x 100m, the picture gets smoother. Focusing on the hotspots when making a decision on the 100 x 100m scale is not very efficient.

If the action level is 25 pCi/g (the top of the color bar), quite a bit of area in 10x10m or 20x20m blocks must be remediated, but if the decision scale is 100x100m hardly any remediation needs to be done.

The action level is also a function of scale. Does it apply to the teaspoon, the bulldozer, or the subdivision?

There is no one answer to this. Scaling is something you have to keep in mind when you're doing sampling, and it should be worked out with the regulators and the stakeholders.

Smart Sampling

Scaling Example Tabulated

Parameters of Up-scaled Distributions

Average Size	Mean	Median	Std. Dev.	Min.	10th centile	90th centile	Max.	n
1x1	6.79	4.33	7.93	0.15	1.27	14.62	158.65	48,400
2x2	6.79	4.79	6.74	0.23	1.56	14.11	81.33	12,100
5x5	6.79	5.17	5.38	0.52	2.01	13.62	47.82	1936
10x10	6.79	5.63	4.29	0.89	2.56	12.37	24.25	484

Scaling is a smoothing process. As we average over bigger and bigger block sizes, the mean value does not change but the standard deviation goes down. There is not as much variability as we smooth towards the mean. The minimum value increases and the maximum value decreases.

Smart Sampling

Decision Example

- *Given an action level of 15ppm, how much needs to be remediated? (Assume 100% reliability)*
- *The answer depends on the scale of decision*

Discretization (in meters)	Proportion	Number of Panels	Total Area of Remediation (m^2)
1x1	0.0956	4629	462,900
2x2	0.0861	1042	416,800
5x5	0.0754	146	365,000
10x10	0.0640	31	310,000

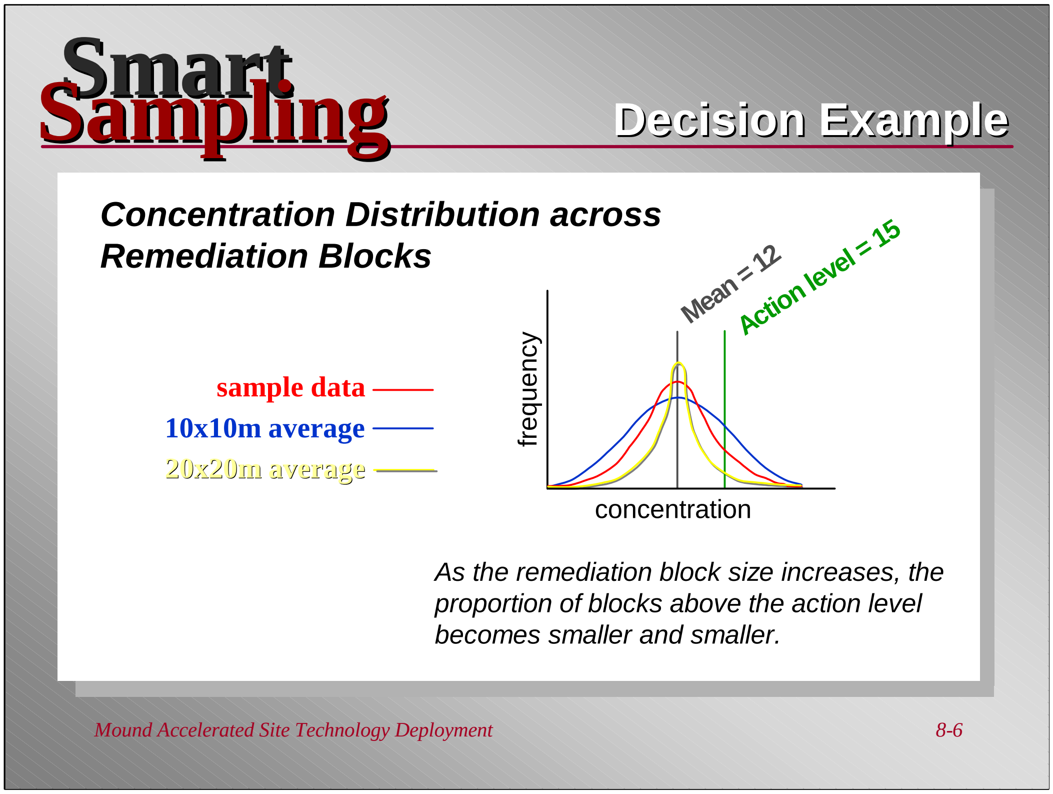
(site is 2.2 x 2.2 km)

Mound Accelerated Site Technology Deployment *8-5*

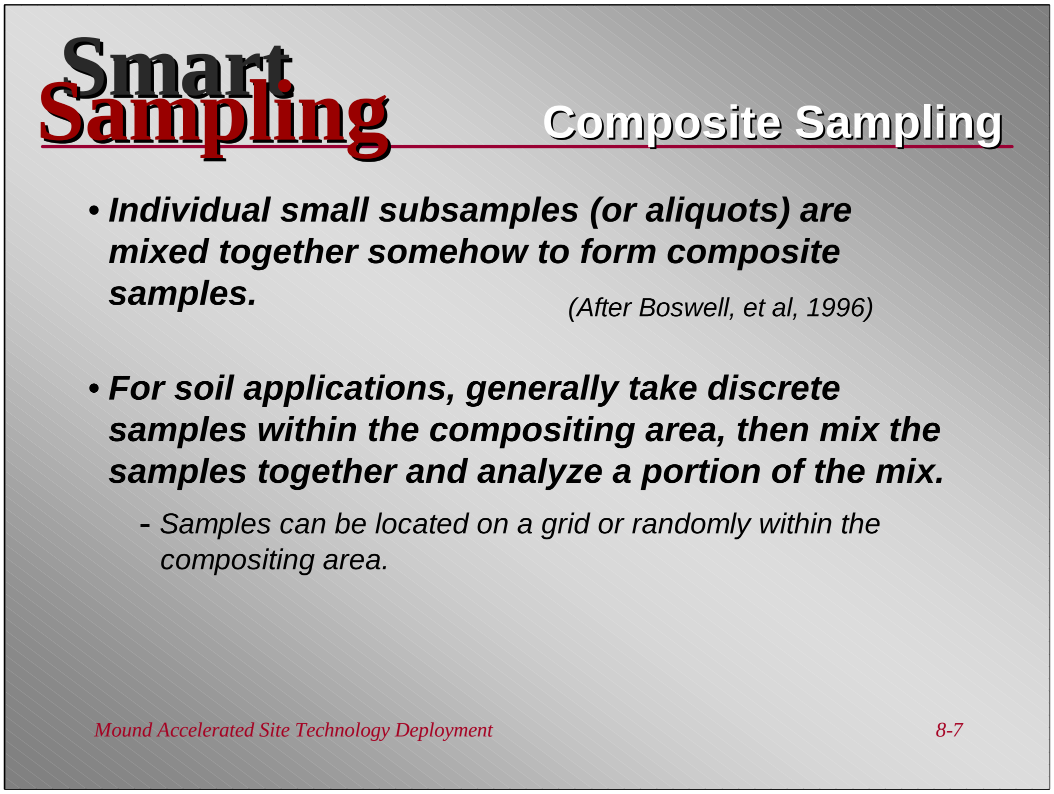
The proportion of the site that requires clean-up drops as the discretization size increases.

The proportion of the distribution that is above 15 ppm is dropping because the action level is above the mean. As we average up, things tend to collapse to the mean (the average of the whole domain is the mean value).

Remediation panel size and action level are fundamentally linked.



If the action level is the mean, the proportions don't change much; and if the action level is below the mean, the proportion of remediation blocks above the action level will increase as the scale increases.



The scale of the compositing area should be agreed upon with the regulator so that their confirmation sampling establishes any remaining concentrations over the same scale.

Smart Sampling

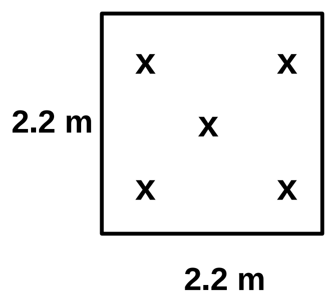
Composite Example

- ***Consider exhaustive data field in scaling example as a 2.2 x 2.2 m (not km) block***
- ***Take 5 samples.***

Sample #	X coord. (cm)	Y coord. (cm)	Value (ppm)
1	110	110	22.78
2	20	20	1.41
3	20	200	0.31
4	200	20	13.38
5	200	200	0.752

(Mean = 7.73 ppm)

Mound Accelerated Site Technology Deployment *8-8*



After mixing the concentration values of the 5 samples together, we get a composite value of 7.73 ppm. So even though there is a hot spot, it is just one of the 5 samples contributing to the average.

If applying the action level of 15 ppm, the composite sample says this 2.2 meter by 2.2 meter block is clean.

In making scaling decisions, you and the regulator consider the contaminant of concern and how the dose is calculated. In most cases, dose is gauged not by peak concentrations, but by the average that a person comes in contact with over some period of time.

Smart Sampling

How to Scale

- ***Ideally, samples could be obtained on the same support as the decisions being made.***
  (e.g., composite samples covering same scale as remediation panel)
- ***In the real world, this rarely happens***
  (either poor understanding or not knowing up front what scale the decision will be made upon, or it's just not feasible to go out and do the compositing)
- ***Explore the following scaling options:***
  - ***Kriging or simulating and then averaging***
  - ***Block Kriging***

Smart Sampling

Block Kriging

- ***For Block Kriging, the kriging system is redefined to estimate the average value within a local area***

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} & 1 \\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} & C_{26} & 1 \\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} & C_{36} & 1 \\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} & C_{46} & 1 \\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} & C_{56} & 1 \\ C_{61} & C_{62} & C_{63} & C_{64} & C_{65} & C_{66} & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \\ \omega_5 \\ \omega_6 \\ \mu \end{bmatrix} = \begin{bmatrix} C_{10} \\ C_{20} \\ C_{30} \\ C_{40} \\ C_{50} \\ C_{60} \\ 1 \end{bmatrix}$$

- $\vec{\mathbf{D}}$ *now contains the covariances between sample points and the estimation area.*
- $\vec{\mathbf{D}}$ *covariances are the average covariances between the sample location and all points within the estimation area.*

- *C matrix stays the same as it provides information on point to point covariances.*

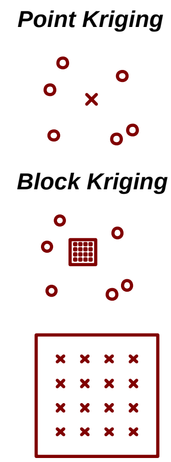
Smart Sampling

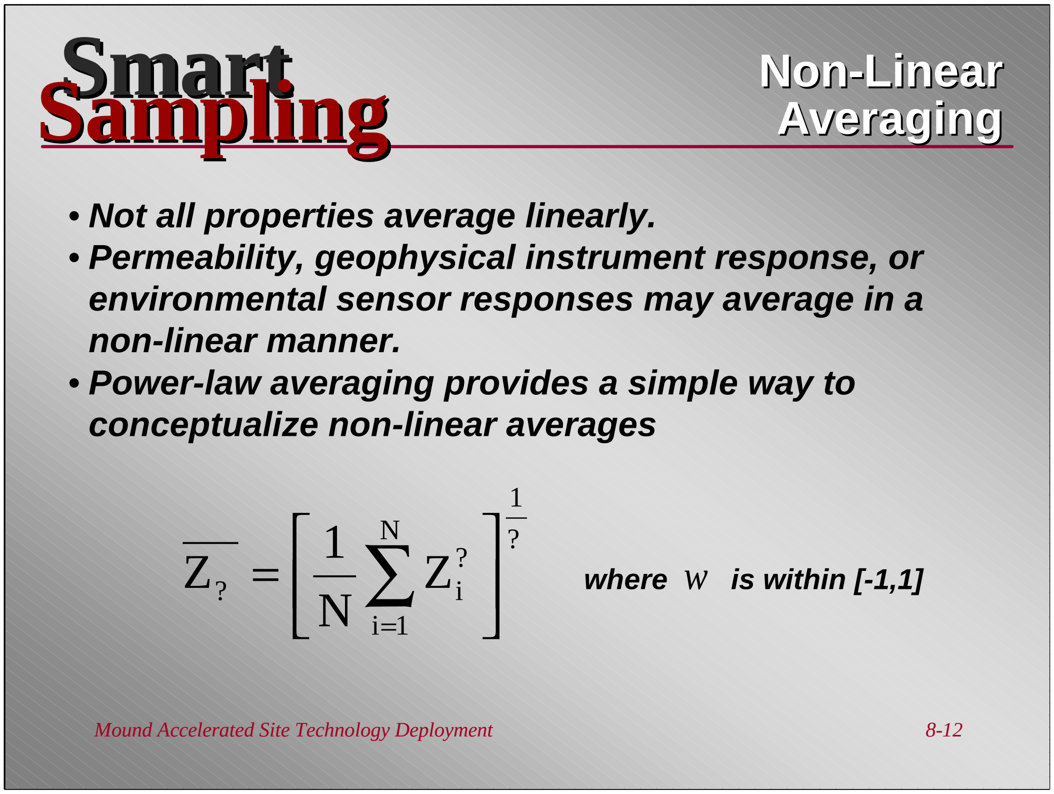
Block Kriging Mechanics

- ***Average covariance between a point and all points within a block is approximated by assuming a finite number of points within a block***
- ***Calculation of this average covariance for n points and then solving one kriging system is computationally much more efficient than solving n kriging systems and then averaging***
- ***How many points are enough?***

(Rule of thumb: 4x4 discretization for 2-D blocks)

(After Isaaks and Srivastava, 1989)





ω = 1: arithmetic averaging (sum up values, divide by number of values)

ω = -1: harmonic average (the inverse of the arithmetic average)

ω = 0: equation blows up; geometric average

Non-linear Averaging Example

Driving 30 miles home from work, what is our average speed?.

10 mph	30 mph	60 mph	Average velocity?
10 mi	10 mi	10 mi	= 30 mi
1 hr	20 min	10 min	= 1.5 hrs

- $$\bar{v}(\text{arithmetic}) = \frac{10 \cdot 10 + 30 \cdot 10 + 60 \cdot 10}{30\text{mi}} = 33.3mph$$

But 30 mi / 1.5 hrs = 20 mph

- $$\bar{v}(\text{harmonic}) = \frac{30}{\left(\frac{10}{10}\right) + \left(\frac{10}{30}\right) + \left(\frac{10}{60}\right)} = 20\text{mph}$$

Smart Sampling

Final Goal

We want to avoid the results of the potato chip analogy.

Samples are like potato chips. You're never satisfied with just one and you never know when you've had enough until you've had too many.

(After J.C. Myers, 1997)